Backup Policy and Backup

Group-5: Team-1

Jay Sharma - **2018101033**
Tanmay Garg - **2018102021**
Vedant Mundheda - **2018112006**

Aim

- Backup and recover datasets stored on the system.
- Maintain version history to access previous versions of the datasets.
- Store only the modifications to datasets (between successive versions) to keep the size small.
- Be able to handle all file formats and directory structures.
- Easy and intuitive backup and recovery API calls should be available.

Extended Aim

Explore how to backup:

- MySQL databases
- MINIO Objects
- Label-Studio Annotations

Base Backup Functionality

```
make_backup(backup_path, dataset_path, full=False)
```

- For backing up a dataset, simply call the “make_backup” function.
- This function takes care of incremental backups, maintaining version history, and storing only those files which are needed to save space.

```
recover_backup(backup_path, dataset_path, out_folder, backup_id=None, req_time=None)
```

- For restoring a dataset using its backup, just call the “recover_backup” function.
- This takes care of how to backup from various full/partial backups, as the dataset was before given a time.

MySQL Database Backup

- MySQL databases are stored in form of IBD (InnoDB engine files) by default. [Database folders containing an IBD file for each table].
- If you lose these files due to corruption, they can be restored. However, the database can't be restored DROP TABLE, TRUNCATE TABLE, or DROP DATABASE commands are used.
- Backup the IBD file folder using the make_backup API call. Then, to restore, call recover_backup to the MySQL data folder (default location is "C:\ProgramData\MySQL\MySQL Server 8.0\Data").
- To make MySQL recognise those files, open the MySQL shell,
    - type "USE db_name", and for each table type:
    - ALTER TABLE table_name DISCARD TABLESPACE
    - ALTER TABLE table_name IMPORT TABLESPACE

REF: https://dev.mysql.com/doc/mysql-enterprise-backup/3.11/en/partial.restoring.single.html

MINIO Object Backup

- If the data is stored in a distributed manner, it would be tough to back it up with this, consider using backup utilities catered to MINIO. [use “mc mirror” command or a utility like Restic/Rclone]
- If the data is only local and data is stored in a MINIO data bucket, then the functions “make_backup” and “recover_backup” can be called at the location where the bucket is stored. (the location is mount_name/bucket_name).
- For example, if you started MINIO server by typing “minio server /tmp”, and created a bucket named “jarvis”, then the data can be found in (and be backed up from) the path “/tmp/jarvis”.

REF:
https://serverfault.com/questions/850998/how-to-keep-minio-backed-up
https://github.com/minio/minio/issues/4135

Label-Studio Annotations Backup

- The same functions “make_backup” and “recover_backup” can be called in whatever location the data is exported to.
- The data can be exported using “Export” feature in Label-Studio. A path is specified to export the data to, same path can be used for backup.
- If the annotations are not complete, use the “Create New Snapshot” feature in Label-Studio. A path is specified to store the snapshot, same path can be used for backup.

REF: https://labelstud.io/guide/export.html

A few things to mention

- Binary files can’t be backed up by diff-check methods, but the files which can be backed up this way (text files), generally don’t get too big to employ diff-check.
- Currently the basic backup API provides backup functionality if it has to be done on the same machine as the one storing the dataset. However, we can add backup from a different machine by employing SSH, FTP or HTTP methods, based on whatever is specified.
- MySQL, MINIO, and Label-Studio backups can be made by the current system.

Contributions

We had an **equal contribution** from each team member. The exact breakdown is given below:

- Jay Sharma
  - API:make_backup
  - MySQL backup
  - Documentation
- Tanmay Garg
  - API:recover_backup
  - MINIO backup
  - Exploring third party backup-tools
- Vedant Mundheda
  - Testing API
  - Documentation
  - Label-Studio Backup

Thank You